6156517

```
  1  GAATTCGCGGCCGCTCGGTTGGTCCTC ATG GTG TCT TTT TGA AGAGGACCTGAGCCTTTCACCCAAATATA  74
  1                              M   V   S   F   *                                    5

 75  TCAAAAAACCCGGCAACCGGCCAAAAAAATTGCAAAGACCTCTCGTAGGCACAAAGACCTATTCTAGCCATCAACTTT  154

155  GTATCCGACGCTGCCGTTTAGCTGCCGTCTTGAAGTCAAGC ATG GCG ACT ACT GAG TCC TCG GCC CCG  223
  1                                            M   A   T   T   E   S   S   A   P      9

224  GCG GCC ACC CAG CCG GCC AGC ACC CCG CTG GCG AAC TCG TCG CTG TAC GTC GGT GAC  283
 10   A   A   T   Q   P   A   S   T   P   L   A   N   S   S   L   Y   V   G   D    29

284  CTG GAG AAG GAT GTC ACC GAG GCC CAG CTG TTC GAG CTC TTC TCC GTT GGC CCT GTG  343
 30   L   E   K   D   V   T   E   A   Q   L   F   E   L   F   S   V   G   P   V    49

344  GCC TCC ATT CGC GTG TGC CGC GAT GCC GTC ACG CGC TCG CTG GGC TAC GCC TAC GTC  403
 50   A   S   I   R   V   C   R   D   A   V   T   R   S   L   G   Y   A   Y   V    69

404  AAC TAC AAC AGC GCT CTG GAC CCC CAG GCT GCT GAC CGC GCC ATG GAG ACC CTG AAC TAC  463
 70   N   Y   N   S   A   L   D   P   Q   A   A   D   R   A   M   E   T   L   N   Y    89

464  CAT GTC GTG AAC GGC AAG CCT ATG CGC ATC ATG TGG TCG CAC CGC GAC CCT TCG GCC CGC  523
 90   H   V   V   N   G   K   P   M   R   I   M   W   S   H   R   D   P   S   A   R   109

524  AAG TCG GGC GTC GGC AAC ATC TTC ATC AAG AAC CTG GAC AAG ACC ATC GAC GCC AAG GCC  583
110   K   S   G   V   G   N   I   F   I   K   N   L   D   K   T   I   D   A   K   A   129

584  CTG CAC GAC ACC TTC TCG GCC TTC GGC AAG ATT CTG TCC TGC AAG GTT GCC ACT GAC GCC  643
130   L   H   D   T   F   S   A   F   G   K   I   L   S   C   K   V   A   T   D   A   149

644  AAC GGC GTG TCG AAG GGC TAC GGC TTC GTG CAC TTC GAG GAC CAG GCC GCT GCC GAT CGC  703
150   N   G   V   S   K   G   Y   G   F   V   H   F   E   D   Q   A   A   A   D   R   169

704  GCC ATT CAG ACC GTC AAC CAG AAG ATT GAG GGC AAG AAG ATC GTG TAC GTG GCC CCC TTC  763
170   A   I   Q   T   V   N   Q   K   I   E   G   K   K   I   V   Y   V   A   P   F   189
```

FIG. 1A

```
764  CAG AAG CGC GCT GAC CGC CCC AGG GCA AGG ACG TTG TAC ACC AAC GTG TTC GTC AAG AAC  823
190   Q   K   R   A   D   R   P   R   A   R   T   L   Y   T   N   V   F   V   K   N   209

824  TTG CCG GCC GAC ATC GGC GAC GAG CTG GGC AAG ATG GCC ACC GAG CAC GGC GAG ATC      883
210   L   P   A   D   I   G   D   E   L   G   K   M   A   T   E   H   G   E   I        229

884  ACC AGC GCG GTG GTC ATG AAG GAC AAG GAC GGC AGC GGC TTC ATC AAC                   943
230   T   S   A   V   V   M   K   D   K   D   G   S   G   F   I   N                    249

944  TTC AAG GAC GCC GAG TCG GCG GCC AAG TGC GTG GAG TAC CTG AAC GAG CGC GAG ATG AGC   1003
250   F   K   D   A   E   S   A   A   K   C   V   E   Y   L   N   E   R   E   M   S    269

1004 GGC AAG ACC CTG TAC GCC CAG CGC GCC GGC AAG ACC GAG CGC GAG ATG CTG              1063
270   G   K   T   L   Y   A   Q   R   A   G   K   T   E   R   E   M   L               289

1064 CAG AAG GCC GAG GAG AGC AAG CAG GAG CGT TAC CTG AAG TAC CAG AGC ATG AAC CTG CGC   1123
290   Q   K   A   E   E   S   K   Q   E   R   Y   L   K   Y   Q   S   M   N   L   R    309

1124 GTC AAG AAC CTG TCC GAC GAG GTC GAC GAC GCC CTG CGT GAG CTG TTC GCC AAC           1183
310   V   K   N   L   S   D   E   V   D   D   A   L   R   E   L   F   A   N            329

1184 TCT GGC ACC ATC ACC TCG TGC AAG GTC ATG AAG GAC GGC AGC GGC AAG TCC GGC TTC       1243
330   S   G   T   I   T   S   C   K   V   M   K   D   G   S   G   K   S   G   F        349

1244 GGC TTC GTG TGC TTC ACC GAC CAC AGC CCC CTG ACC CGG GCC ACC GAG ATG GAG ATG AAC   1303
350   G   F   V   C   F   T   D   H   S   P   L   T   R   P   P   V   T   E   M   N    369

1304 GGC AAG ATG GTC AAG GGC AAG CCC CTG TAC GTG GCC CTG GCG CAG CGC AAG GAC GTG CGC   1363
370   G   K   M   V   K   G   K   P   L   Y   V   A   L   A   Q   R   K   D   V   R    389

1364 CGT GCC ACC CAG CTG GAG GCC AAC ATG CAG GCG CGC ATG GGC GCC ATG AGC CGC           1423
390   R   A   T   Q   L   E   A   N   M   Q   A   R   M   G   A   M   S   R            409
```

FIG. 1B

```
1424  CCG CCG AAC CCG ATG GCC GGC ATG AGC CCC TAC CCC GGC GCC ATG CCG TTC TTC GCT CCC  1483
 410   P   P   N   P   M   A   G   M   S   P   Y   P   G   A   M   P   F   F   A   P    429

1484  GGC CCC GGC GGC ATG GCT GCT CCG CGC ATG TAC CCC ATG ATG TTC ATG TAC CCC ATG ATG  1543
 430   G   P   G   G   M   A   A   P   R   M   Y   P   M   M   F   M   Y   P   M   M    449

1544  CCG CCG CGC GGC CCT GGC ATG CCC ATG GGC CCC ATG GGC CCC ATG ATG GGC CCC CCC CAG  1603
 450   P   P   R   G   P   G   M   P   M   G   P   M   G   P   M   M   G   P   P   Q    469

1604  ATG ATG GGT GGC CCC ATG GGC CCC ATG ATG GGC CGC GGC CAG GGC CGC GGC CCC GGC CGC  1663
 470   M   M   G   G   P   M   G   P   M   M   G   R   G   Q   G   R   G   P   G   R    489

1664  GGC CCC TCC GGC CAG GGC CGC GGC AAC AAC GCC CCT GCC CAG CCG GCC CAG CAG AAG CCC  1723
 490   G   P   S   G   Q   G   R   G   N   N   A   P   A   Q   P   A   Q   Q   K   P    509

1724  GCC GCT GAG CCG GCC GCC CCG GCC GCC GCT CCC GCC GCC CAG CCG GCC GCC GCG GCC GCC  1783
 510   A   A   E   P   A   A   P   A   A   A   P   A   A   Q   P   A   A   A   A   A    529

1784  GCG GCG GAG CCG GCC GCC CAG CAG CAG CCG CTG ACC GCC TCC GCG CTG GCC CCT GCC GCC  1843
 530   A   A   E   P   A   A   Q   Q   Q   P   L   T   A   S   A   L   A   P   A   A    549

1844  GCC GCG GAG CAG GAG CAG AAG ATG ATG ATC GGC GAG CGC CTG TAC CCG CAG GTG GCG GAG  1903
 550   A   A   E   Q   E   Q   K   M   M   I   G   E   R   L   Y   P   Q   V   A   E    569

1904  CTG CAG CCC GAC CTG GCT GGC AAG ATC ACC GGC ATG CTG CTG GAG ATG GAC AAC GCC GAG  1963
 570   L   Q   P   D   L   A   G   K   I   T   G   M   L   L   E   M   D   N   A   E    589

1964  CTT CTG ATG CTT CTG GAG TCG CAC GAG GCG CTG GTG TCC AAG GTG GAC GAG GCC ATC GCT  2023
 590   L   L   M   L   L   E   S   H   E   A   L   V   S   K   V   D   E   A   I   A    609

2024  GTG CTC AAG CAG CAC AAC GTG ATT GCC GAG GAG AAC AAG GCT TAA AGGGCCTGCACGCTGTGCG  2088
 610   V   L   K   Q   H   N   V   I   A   E   E   N   K   A   *                        624
```

FIG. 1C

```
2089 GGCTGGTGGCGGCCGGCGGCGGCGCTGGCGCTGCTTGGGCCGCCGGCAGC ATG GGC GCG GCG GAC GCG GCG GTG TGG 2159
1                                                      M   G   A   A   D   A   A   V   W    8

2160 GAG CAG TGC TTG CTG CTT CTG GCC GCC GTG AAG CCG CGC CGA ACT GGG GCG GAC GGC AGG 2219
9    E   Q   C   L   L   L   L   A   A   V   K   P   R   R   T   G   A   D   G   R   28

2220 CTG GCG TTG ACG CCG GCG CGC CAC AAC ACA AAG TTG GTG GCG TGA AAGTCTCTGGGCGTCCG 2284
29   L   A   L   T   P   A   R   H   N   T   K   L   V   A   *                      43

2285 GACGGGTTGTAAGGTTTAAGAACTGGCTTTTGGCCGGGTTGCGCCCAAAGGCGGAACGGCGGTCTTTTCAGGCCAATCA 2364

2365 CATCCGGCTGAAAAATTCTTACCAAAGCCAACCCCTGCACCCAAAAATTTCGGGGTTCCGAAAGAACTCCCCTTTTT 2444

2445 CCGGCAACGCGTTCTTCAGGCCAATCACTTCCGGGTTGGAAGAAA ATG TTA CCC GGA AAA GGC GGG AAG 2516
1                                                M   L   P   G   K   G   G   K    8

2517 CCC CCT GCA CCC GGA CAA GTT ATT CGG GGT TTC GCC GGG AAT GAG CAA GCG TTC GGG CTG 2576
9    P   P   A   P   G   Q   V   I   R   G   F   A   G   N   E   Q   A   F   G   L   28

2577 TTG GCC GTA TCG CGA ACG CTG TCG GGG TGT CAG GCG CCA GAA GAG AGG ATG ACG TTT TGG 2636
29   L   A   V   S   R   T   L   S   G   C   Q   A   P   E   G   R   M   T   F   W   48

2637 TGA AGGGGTGCAAACTGAGCACACGAGTTTTGGCAATAGAGCTGTGAGGCGGATAGCGGA 2715
49   *                                                          49

2716 ATCAAGCGTGGCGGTCCCTGGCGAGACGAGAGCGCTTCTGTGTTTTGCTGAGCCCTTG ATG GCA CAA TCG CAC 2790
1                                                              M   A   Q   S   H    5

2791 TGT TTT GAG CAG GCG ACT GTA AAG TGC CCG ACG CTA AAA AAG CGG CCG CGA ATT CC 2846
6    C   F   E   Q   A   T   V   K   C   P   T   L   K   K   R   P   R   I       23
```

FIG. 1D

MNRWNLLALTLGLLLVAAPFTKHQFAHASDEYEDDEEDDAPAAP

KDDDVDVTVTVKNWDETVKKSKFALVEFYAPWCGGHCKTLKPEYAKAATALKAAAPDA

LIAKVDATQEESLAQKFGVQGYPTLKWFVDGELASDYNGPRDADGIVGWVKKKTGPPA

VTVEDADKLKSLEADAEVVVVGYFKALEGEIYDTFKSYAAKTEDVVFVQTTSADVAKA

AGLDAVDTVSVVKNFAGEDRATAVLATDIDTDSLTAFVKSEKMPPTIEFNQKNSDKIF

NSGINKQLILWTTADDLKADAEIMTVFREASKKFKGQLVFVTVNNEGDGADPVTNFFG

LKGATSPVLLGFFMEKNKKFRMEGEFTADNVAKFAESVVDGTAQAVLKSEAIPEDPYE

DGVYKIVGKTVESVVLDETKDVLLEVYAPWCGGHCKKLEPIYKKLAKRFKKVDSVIIAK

MDGTENEHPEIEVKGFPTILFYPAGSDRTPIVFEGGDRSLKSLTKFIKTNAKIPYELP
KKGSDGDEGTSDDKDKPASDKDEL

```
  1 gagtacgttt acgccatgaa ccgttggaac cttcttgccc ttaccctggg gctgctgctg
 61 gtggcagcgc cctttaccaa gcaccagttt gctcatgctt ccgatgagta tgaggacgac
121 gaggaggacg atgcccccgc cgcccctaag gacgacgacg tcgacgttac tgtggtgacc
181 gtcaagaact gggatgagac tgtcaagaag tccaagttcg cgcttgtgga gttctacgct
241 ccttggtgcg gccactgcaa gaccctcaag cctgagtacg ctaaggctgc caccgcccctg
301 aaggctgctg ctcccgatgc ccttatcgcc aaggtcgacg ccacccagga ggagtccctg
361 gcccagaagt tcggcgtgca gggctacccc ccgcgacgct ggttcgttga tggcgagctg
421 gcttctgact acaacggccc ccgcgacgct gatggcattg ttggctgggt gaagaagaag
481 actggccccc ccgccgtgac cgttgaggac gccgacaagc tgaagtccct ggaggcggac
541 gctgaggtcg ttgtcgtcgg ctacttcaag gccctggagg gcgagatcta cgacaccttc
601 aagtcctacg ccgccaagac cgaggacgtg gtgttcgtgc agaccaccag cgccgacgtc
```

FIG. 2A

5

```
 661  gccaaggccg  ccggcctgga  cgccgtggac  accgtgtccg  tggtcaagaa  cttcgccggt
 721  gaggaccgcg  ccaccgccgt  cctggccacg  gacatcgaca  ctgactccct  gaccgcgttc
 781  gtcaagtcgg  agaagatgcc  ccccaccatt  gagttcaaac  agaagaactc  tgacaagatc
 841  ttcaacagcg  gcatcaacaa  gcagctgatt  ctgtggacca  ccgccgacga  cctgaaggcc
 901  gacgccgaga  tcatgactgt  gttccgcgag  gccagcaaga  agttcaaggg  ccagctggtg
 961  ttcgtgaccg  tcaacaacga  gggcgacggc  gccgaccccg  tcaccaactt  cttcggcctc
1021  aagggcgcca  cctcgccctgt gctgctgggc  ttcttcatgg  agaagaacaa  gaagttccgc
1081  atggaggggcg agttcacggc  tgacaacgtg  gctaagttcg  ccgagagcgt  ggtggacggc
1141  accgcgcagg  ccgtgctcaa  gtcggaggcc  atccccgagg  acccctatga  ggatggcgtc
1201  tacaagattg  tgggcaagac  cgtggagtct  gtggttctgg  acgagaccaa  ggacgtgctg
1261  ctggaggtgt  acgccccctg  gtgcgcggcac tgcaagaagc  tggagcccat  ctacaagaag
1321  ctggccaagc  gctttaagaa  ggtggattcc  gtcatcatcg  ccaagatgga  tggcactgag
1381  aacgagcacc  ccgagatcga  ggtcaagggc  ttccctacca  tcctgttcta  tcccgccggc
1441  agcgaccgca  cccccatcgt  gttcgagggc  ggcgaccgct  cgctcaagtc  cctgaccaag
1501  ttcatcaaga  ccaacgccaa  gatcccgtac  tgcaagaagc  agaaggctc   cgacggcgac
1561  gaggcacct   cggacgacaa  ggacaagccc  gcgtccgaca  aggacgagct  gtaagcggct
1621  atctgaacta  ccccaggttt  ggagcgtctg  cttgcgcgct  tgcgcttgca  cactgtgcat
1681  ggatggggagt taaggaggag  acggagcacg  gaggctgcgc  tcggttggtg  gcttggagca
1741  ccggcagcgc  gtgatccgtc  ctggcagcag  caacggcgga  gcgggcgcat  attggcgcga
1801  gctggcgagc  ggctgtttgct ggagaggata  tgctgccggg  cgggaggaag  ggctagggc
1861  agagatgaga  gcgttacggg  ctggcatgcg  ggcgccgtg   cctctccctg  cggtgcagtc
1921  cttgctagga  gacgcacggt  tttgccaaag  agggacgctg  tccacagccc  tgcgactgga
1981  agtttttag   gccctgcggt  ggtagtggtg  ttggtacggt  tgtgtgcata  agatgaacaa
2041  cgtttctctc  aagacgagac  tactagtatg  ctgacggtgt  gtgtatgtgg  tggatggatt
2101  gtgccccgac  catgaagagt  gctgtgttgc  ctcggcgctt  ctgtcgccct  ggatgtgcgt
2161  ggttccgaac  gctggagtca  tctgttgagg  agcgagggtg  ttgtcgggtc  cgcccggcac
2221  ggccgcgtga  tgtccggatg  gggattgcga  ggatgatgag  ccgcagcgca  gatagcgccg
2281  cagcggatcg  agctagcgca  ggatgatgag  agccggggcct tcgcggcgtg  ggatcagggga
2341  ggagccaagg  cggagtgcat  gcgaggaaaa  cagtgtgcgg  caaagaacgg  gctgcaagaa
2401  cgccttgcgc  aaa
```

FIG. 2B

FIG. 3A

```
                31                    A        A              B
                                              -250  ·  ·  -200
CGTCCTATTTTAATACTCCGAAGGAGGCAGTTGGCAGGCAACTGCCACTGACGTC

                      B        31                  C
CCGTAAGGGTAAGGGGACGTCCCACTGGCCGTCCCGTAAGGGGACGTAGTTAGTGCTAGTTAGGTAACTAAC
                                        -150|

              C                 -100      D          -50
GTTTGATTTTTTGTG G[TATAAT]ATATGTACCATGCTTTTAATAGAAGCTTGAATTTATAAATTAAAATATTT[TACAAT]
                        S1

                 S1
ATTTTAC[GGAGAAATT]AAAAAAATTAACAT ATG ACA GCA ATT TTA GAA CGT CGT GAA AAT   10
                                 Met Thr Ala Ile Leu Glu Arg Arg Glu Asn
                                                                      Ser

TCT AGC CTA TGG GCT CGT TTT TGT GAG TGG ATC ACT TCA ACT GAA AAC CGT TTA TAC ATC
Ser Ser Leu Trp Ala Arg Phe Cys Glu Trp Ile Thr Ser Thr Glu Asn Arg Leu Tyr Ile   30
Glu                                                                  Gly
                                                                     Asn
                             .50                                          .150
GGT TGG TTC GGT GTA ATC ATG ATC CCA TGT CTT CTT ACT GCA ACA TCA GTA TTC ATC ATC
Gly Trp Phe Gly Val Ile Met Ile Pro Cys Leu Leu Thr Ala Thr Ser Val Phe Ile Ile   50
                             .100             Leu                             Thr
                                                                     -200
GCT TTC ATC GCT GCT CCG CCA GTA GAC ATC [GAT GGT ATC CGT GAA CCA GTT TCA GGT TCT
Ala Phe Ile Ala Ala Pro Pro Val Asp Ile][Asp Gly Ile Arg Glu Pro Val Ser Gly Ser   70
```
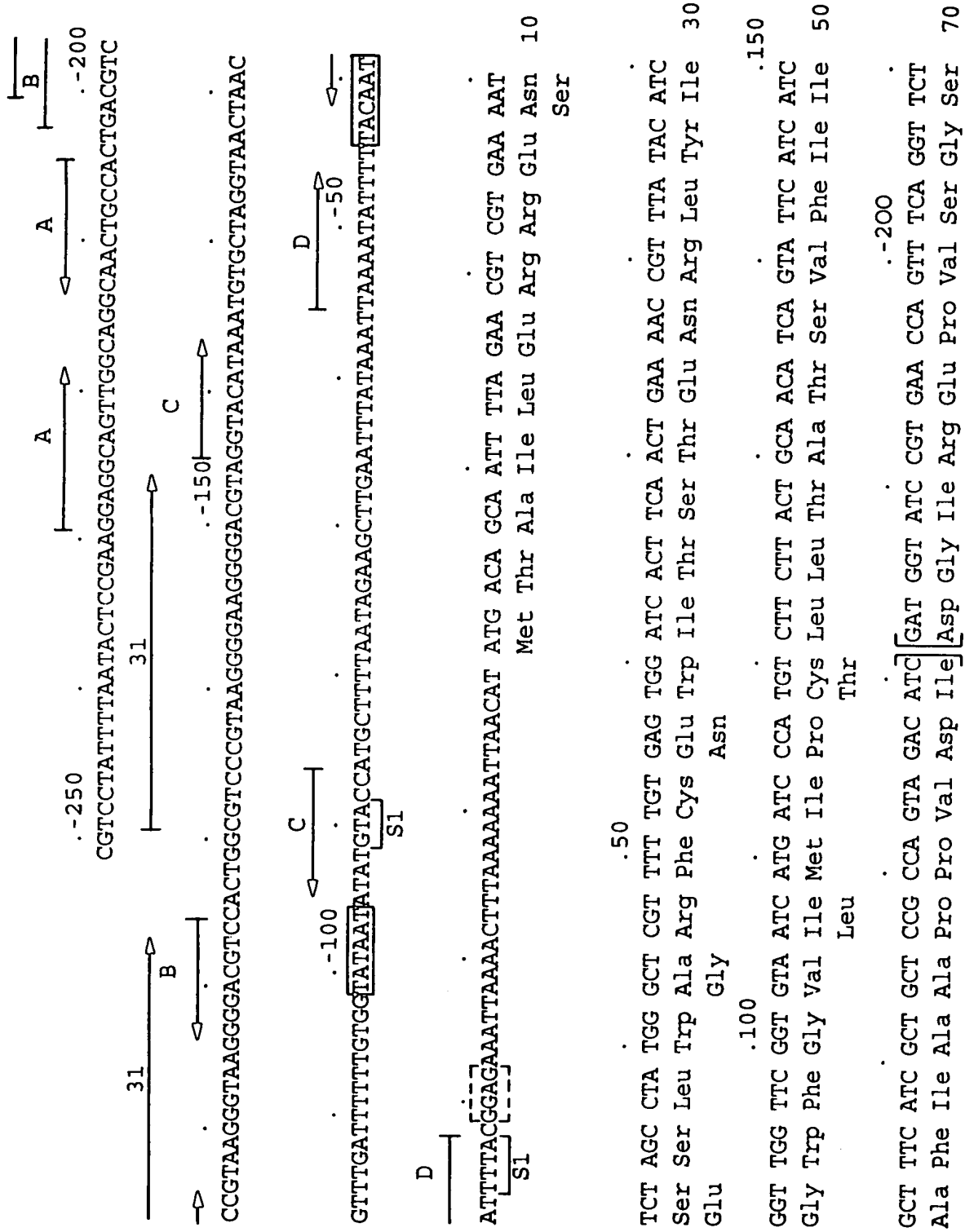
CTT CTT TAC GGT AAC AAC ATC ATT ACA GGT GCT GTA ATC CCA ACT TCT AAC GCA ATC GGT
Leu Leu Tyr Gly Asn Asn Ile Ile Thr Gly Ala Val Ile Pro Thr Ser Asn Ala Ile Gly 90
                                  Ser   Ala

CTT CAC TTC TAC CCA ATT TGG GAA GCT TCT CTA GAC GAG TGG TTA TAC AAC GGT GGT
Leu His Phe Tyr Pro Ile Trp Glu Ala Ser Leu Asp Glu Trp Leu Try Asn Gly Gly 110
                                                                 Val

CCT TAC CAA CTT ATC GTT TGT CAC TTC CTT CTA GGT GTA TAC TGC TAC ATG GGT] [CGT GAG
Pro Tyr Gln Leu Ile Val Cys His Phe Leu Leu Gly Val Tyr Cys Tyr Met Gly] [Arg Glu 130
                                    Glu   Ala

TGG GAA TTA TCT TTC CGT ATG CCA TGG ATC GCT TAC TCA GCT CCA
Trp Glu Leu Ser Phe Arg Met Gly Trp Ile Ala Tyr Ser Ala Pro 150

GTA GCT GCA TCA GCT GTA TTC TTA GTT TAC CCT ATC GGC CAA GGT TCA TTC TCT GAC
Val Ala Ala Ser Ala Val Phe Leu Val Tyr Pro Ile Gly Gln Gly Ser Phe Ser Asp 170
                                        Ile

GGT ATG CCT TTA GGT] [ATC TCT GGT ACT TTC AAC ATC GTA TTC CAA GCA GAA CAC
Gly Met Pro Leu Gly] [Ile Ser Gly Thr Phe Asn Ile Val Phe Gln Ala Glu His 190

AAC ATC CTT ATG CAC CCA TTC CAC ATG TTA GGT GTT GCT GGT GTA TTC GGT TCA TTA
Asn Ile Leu Met His Pro Phe His Met Leu Gly Val Ala Gly Val Phe Gly Gly Ser Leu 210

TTC TCA GCT ATG CAC GGT TCT TTA GTT ACT TCA TCT TTA GTT ATC CGT GAA ACA ACT GAA AAC
Phe Ser Ala Met His Gly Ser Leu Val Thr Ser Ser Leu Val Ile Arg Glu Thr Thr Glu Asn 230

FIG. 3B

6

```
                                                                          .700
GAA TCA GCT AAC GAA GGT TAC CGT TTC GGT CAA GAA GAA ACT TAC AAC ATT GTA GCT
Glu Ser Ala Asn Glu Gly Tyr Arg Phe Gly Gln Glu Glu Thr Tyr Asn Ile Val Ala 250
                                                                          .750

GCT CAT][GGT TAC TTT GGT CGT CTA ATC TTC CAA TAC GCT TCT TTC AAC AAC TCT CGT TCA
Ala His] Gly Tyr Phe Gly Arg Leu Ile Phe Gln Tyr Ala Ser Phe Asn Asn Ser Arg Ser 270
                                                                          .800

TTA CAC TTC TTC TTA GCT GCT TGG CCG GTA ATC GGT ATT TGG TTC ACT GCT TTA GGT TTA
Leu His Phe Phe Leu Ala Ala Trp Pro Val Ile Gly Ile Trp Phe Thr Ala Leu Gly Leu 290
                                      Val                                      Ile
                                                                          .850

TCA ACT ATG GCA TTC AAC TTA AAC GGT TTC AAC CAA TCA GTA GTA GAC TCA CAA
Ser Thr Met Ala Phe Asn Leu Asn Gly Phe Asn Gln Ser Val Val Asp Ser Gln 310
                                                                          .900

GGT CGT GTA CTA AAC ACT TGG GCA GAC ATC ATC AAC CGT GCT AAC TTA GGT ATG GAA GTA
Gly Arg Val Leu Asn Thr Trp Ala Asp Ile Ile Asn Arg Ala Asn Leu Gly Met Glu Val 330
                                          Ile
                                                                          .950

ATG CAC GAG CGT AAC GCT CAC AAC TTC CCT CTA GAC TTA GCT TCA ACT AAC TCT AGC TCA
Met His Glu Arg Asn Ala His Asn Phe Pro Leu Asp Leu Ala Ser Thr Asn Ser Ser Ser 350
                                                                    .1000      .1050
                                                         Ala Ile Glu Glu Ala Pro

AAC AAC TAA TTT TTTTTTAAACTAAAATAAATCTGGTTAACCATACCTAGTTTATTTTAGTTTATACACACTTTT
Asn Asn *Oc                                                                    S1
Thr Gly *Oc
                                                              .1100
                                                                          .1150
CATATATATATACTTAATAGCTACCATAGGCAGTTGGCAGGACGTCCC
```
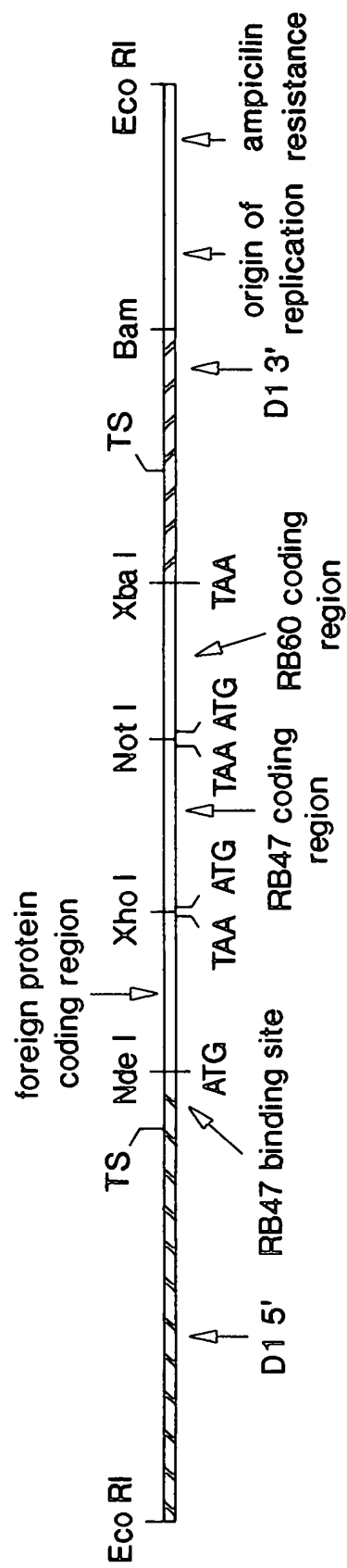
FIG. 3C

FIG. 4

TS = transcription start and transcription stop
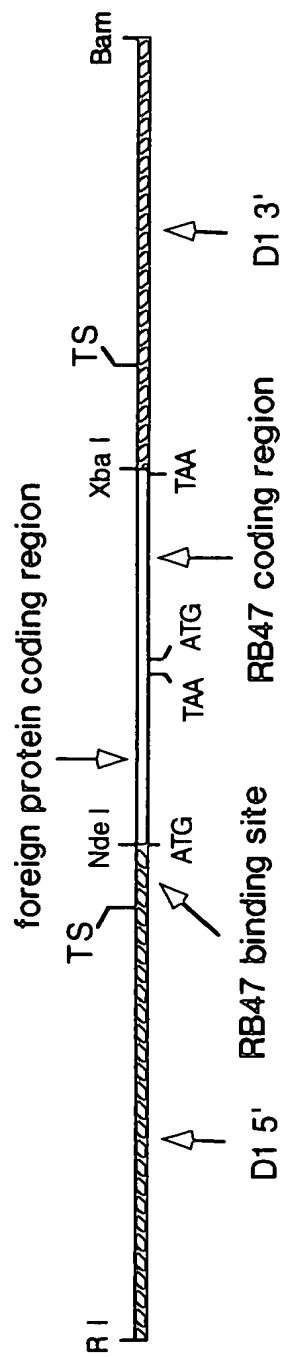
```
  1  ATG GGC CAT CAT CAT CAT CAT CAT AGC AGC GGC CAT ATC GAA GGT CGT         60
  1   M   G   H   H   H   H   H   H   S   S   G   H   I   E   G   R          20

 61  CAT ATG GCG ACT ACT GAG TCC TCG GCC GCG CCG GCC AGC ACC CCG            120
 21   H   M   A   T   T   E   S   S   A   A   P   A   S   T   P             40

121  CTG GCG AAC TCG TCG CTG TAC GTC GGT GAC CTG GAG AAG GAT GCC CAG CTG    180
 41   L   A   N   S   S   L   Y   V   G   D   L   E   K   D   A   Q   L      60

181  TTC GAG CTC TTC TCC TCG GTT GGC CCT GTG GCC TCC ATT CGC GTG TGC GAT GCC GTC   240
 61   F   E   L   F   S   S   V   G   P   V   A   S   I   R   V   C   D   A   V     80

241  ACG CGC CGC TCG CTG GGC TAC TAC AAC AGC AGC GCT CTG GAC CCC CAG GCT    300
 81   T   R   R   S   L   G   Y   Y   N   S   N   A   L   D   P   Q   A      100

301  GCT GAC CGC GCC ATG GAG ACC CTG AAC TAC CAT GTG GTG AAC GGC AAG CCT ATG CGC ATC   360
101   A   D   R   A   M   E   T   L   N   Y   H   V   V   N   G   K   P   M   R   I    120

361  ATG TGG TCG CAC CGC GAC CCT TCG GCC CGC AAG TCG GGC GTC GGC AAC ATC TTC ATC AAG   420
121   M   W   S   H   R   D   P   S   A   R   K   S   G   V   G   N   I   F   I   K    140

421  AAC CTG GAC AAG ACC ATC GAC GCC AAG GCC CTG GCC CAC GAC ACC TTC TCG GCC TTC GGC AAG   480
141   N   L   D   K   T   I   D   A   K   A   L   A   H   D   T   F   S   A   F   G   K    160

481  ATT CTG TCC TGC AAG GTT GCC ACT GAC GCC AAC GGC GTG TCG AAG GGC TAC TTC GTG   540
161   I   L   S   C   K   V   A   T   D   A   N   G   V   S   K   G   Y   F   V    180

541  CAC TTC GAG GAC CAG GCT GCC GCC GAT CGC GCC ATT CAG ACC GTC AAC CAG AAG AAG ATT   600
181   H   F   E   D   Q   A   A   A   D   R   A   I   Q   T   V   N   Q   K   K   I    200

601  GAG GGC AAG ATC GTG TAC GTG GCC CCC TTC CAG AAG CGC GCT GAC CGC GCC CCC AGG GCA AGG   660
201   E   G   K   I   V   Y   V   A   P   F   Q   K   R   A   D   R   A   P   R   A   R    220
```

FIG. 5A

```
661 ACG TTG TAC ACC AAC GTG TTC GTC AAG AAC TTG CCG GCC GAC ATC GGC GAC GAC GAG CTG  720
221  T   L   Y   T   N   V   F   V   K   N   L   P   A   D   I   G   D   D   E   L   240

721 GGC AAG ATG GCC ACC GAG CAC GGC GAG ATC ACC AGC GCG GTG GTC AAG GAC AAG          780
241  G   K   M   A   T   E   H   G   E   I   T   S   A   V   V   K   D   K           260

781 GGC GGC AGC AAG GGC TTC GGC TTC ATC AAC TTC AAG GAC GCG GAG TCG GCG GCC AAG TGC  840
261  G   G   S   K   G   F   G   F   I   N   F   K   D   A   E   S   A   A   K   C   280

841 GTG GAG TAC CTG AAC GAG CGC GAG ATG AGC GGC AAG ACC CTG TAC GCC CGC GCC CAG      900
281  V   E   Y   L   N   E   R   E   M   S   G   K   T   L   Y   A   R   A   Q       300

901 AAG AAG ACC GAG CGC GAG GCG ATG CTG CGC CAG AAG GCC GAG GAG AGC AAG CAG GAG CGT  960
301  K   K   T   E   R   E   A   M   L   R   Q   K   A   E   E   S   K   Q   E   R   320

961 TAC CTG AAG TAC CAG AGC ATG CTG CGT GAG CTG TTC GCC AAC TCT GGC ACC ATC ACC TCG 1020
321  Y   L   K   Y   Q   S   M   L   R   E   L   F   A   N   S   G   T   I   T   S   340

1021 GAC GAC GCC CTG CGT GAG CTG AAG TCC GGC AAG GTC GTG AAG ATG GTC GAG GAG GTC GAC 1080
341   D   D   A   L   R   E   L   K   S   G   K   V   V   K   M   V   E   E   V   D  360

1081 AAG GAC GGC AGC AAG GGC TTC GTG TGC TTC ACC AGC CAC GAC GAG                     1140
361   K   D   G   S   K   G   F   V   C   F   T   S   H   D   E                      380

1141 GCC ACC CGG CCG CCC GTG ACC GAG ATG AAC GGC AAG ATG GTC AAG GGC AAG CCC CTG TAC 1200
381   A   T   R   P   P   V   T   E   M   N   G   K   M   V   K   G   K   P   L   Y  400

1201 GTG GCC CTG GCG CAG CGC CGT GCC ACC CAG CTG GAG GCC AAC ATG CAG                 1260
401   V   A   L   A   Q   R   R   A   T   Q   L   E   A   N   M   Q                  420

1261 GCG CGC ATG TAA GGATCC                                                         1278
421   A   R   M   *                                                                 424
```
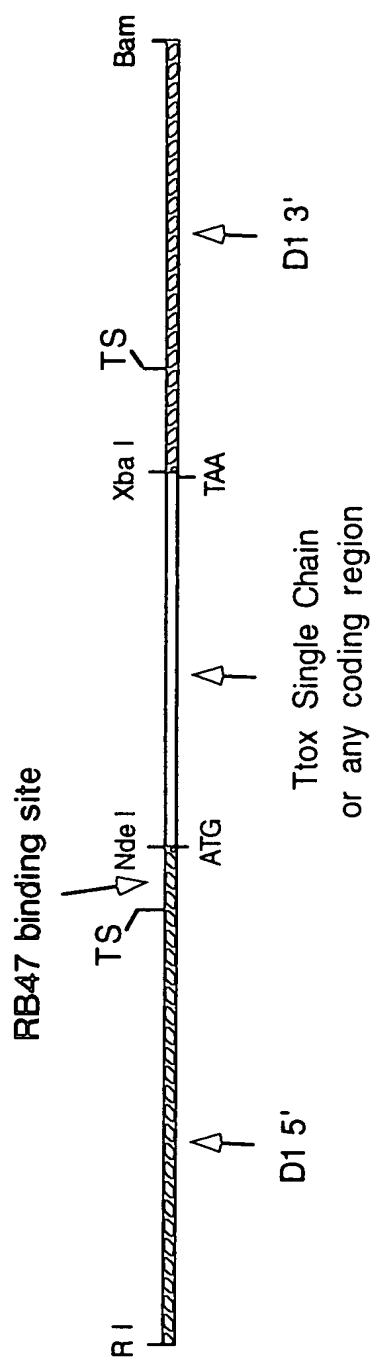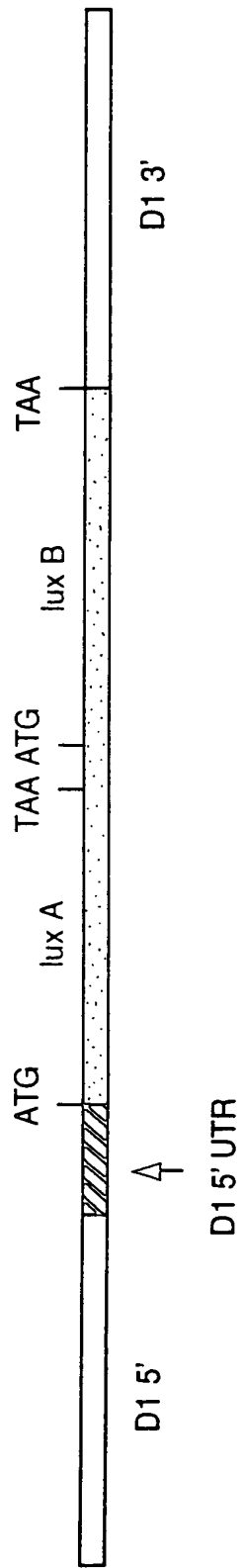
FIG. 5B

foreign protein coding region

TS   Nde I   Xba I   TS
         ATG   TAA ATG   TAA
D1 5'   RB47 binding site   RB47 coding region   D1 3'

TS = transcription start and transcription stop

FIG. 6

TS   Nde I   Xba I   TS
RB47 binding site   ATG   TAA
D1 5'   Ttox Single Chain   D1 3'
         or any coding region

FIG. 7

FIG. 8

Bacterial luciferase A and B proteins expressed from a single mRNA containing the psbA 5' UTR with translational activator element.
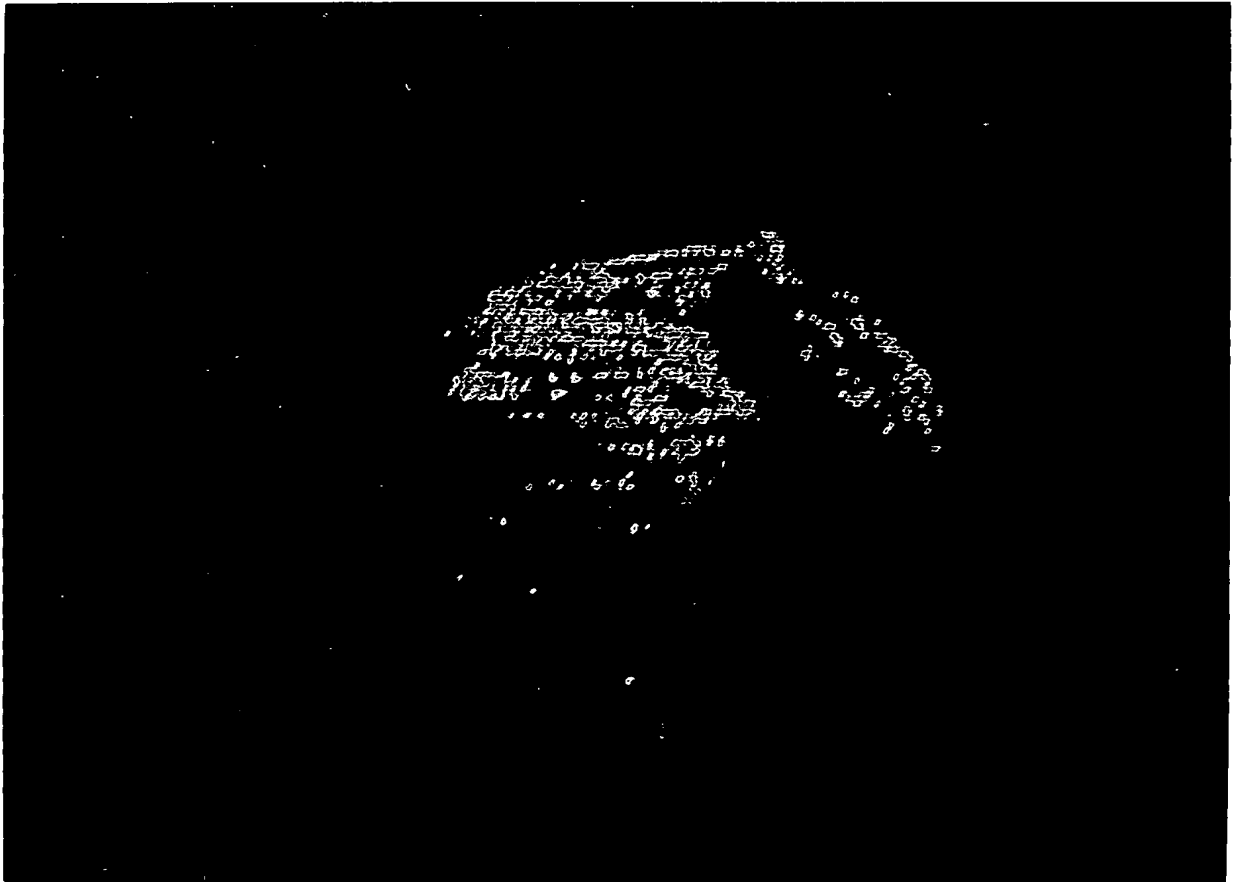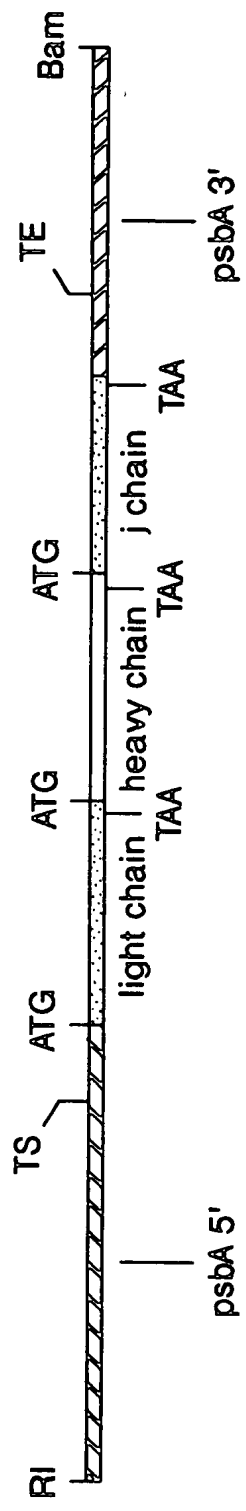
FIG. 9

FIG. 10

FIG. 11